Transcriptome dynamics during fibre development in contrasting genotypes of *Gossypium hirsutum* L.

Deepti Nigam¹,†, Palcharaka Kavita¹,†, Rajiv K. Tripathi¹, Alok Ranjan¹, Ridhi Goel¹,², Mehar Asif¹,², Alpika Shukla¹, Gunjita Singh¹, Debasis Rana³ and Samir V. Sawant¹,²,✉

¹Plant Molecular Biology Laboratory, CSIR-National Botanical Research Institute, Rana Pratap Marg, Lucknow, India
²Academy of Scientific & Innovative Research, New Delhi, India
³JK Agri-Genetics Ltd, Hyderabad, India

Summary

Understanding the contribution of genetic background in fibre quality traits is important for the development of future cotton varieties with superior fibre quality. We used Affymetrix microarray (Santa Clara, CA) and Roche 454 GSFLX (Branford, CT) for comparative transcriptome analysis between two superior and three inferior genotypes at six fibre developmental stages. Microarray-based analysis of variance (ANOVA) for 89 microarrays encompassing five contrasting genotypes and six developmental stages suggests that the stages of the fibre development have a more pronounced effect on the differentially expressed genes (DEGs) than the genetic background of genotypes. Superior genotypes showed enriched activity of cell wall enzymes, such as pectin methyl esterase, at early elongation stage, enriched metabolic activities such as lipid, amino acid and ribosomal protein subunits at peak elongation, and prolonged combinatorial regulation of brassinosteroid and auxin at later stages. Our efforts on transcriptome sequencing were focused on changes in gene expression at 25 DPA. Transcriptome sequencing resulted in the generation of 475 658 and 429 408 high-quality reads from superior and inferior genotypes, respectively. A total of 24 609 novel transcripts were identified from a comparison of contrasting genotypes and six developmental stages. These results indicate the benefits of providing more thorough qualitative and quantitative description of gene expression than the previous microarray-based assay during the early stages (2 to 1 DPA and 2–8 DPA) of cotton fibre development (Wang et al., 2010b).

A previous study using microarray and quantitative gene expression analyses has indicated that ethylene is involved in fibre cell elongation (Shi et al., 2006). It was found that brassinosteroid (BR) promotes fibre cell development on cultured cotton ovules (Sun and Allen, 2005) in a manner similar to the well-established requirement for plant hormones gibberellic acid (GA) and auxin (Beasley et al., 1974). Many candidate genes for hormone biosynthesis or other fibre developmental stages that are expressed in cotton fibre cells have been cloned and characterized (Delmer et al., 1995; Kim and Triplett, 2004; Suo et al., 2003). Recently, a study was carried out for mining of genes that potentially control the improved qualities of domesticated cotton fibre (Rapp et al., 2010). In another excellent and in-depth analysis, Paterson et al. (2012) examined the repeated polyploidization in *Gossypium* spp. up to the level of five- to sixfold along with its macro- and microsyntenic relationships with grape and cacao for understanding the genetic basis of domesticated phenotypes. However, very few studies understood the impact of genetic background that leads to multiple genotypes with

Introduction

Cotton fibres are single-cell seed trichomes that originate from epidermal cells of ovule at or just before anthesis (Basra and Malik, 1984). The fibre development takes place in four overlapping but distinct stages: initiation, elongation, secondary cell wall (SCW) synthesis and maturation (Basra and Malik, 1984; Wilkins and Jernstedt, 1999). Each of the developmental stages in cotton fibre has a distinct gene expression profile (Hinchliffe et al., 2010; Lee et al., 2007; Shi et al., 2006; Smart et al., 1998). Despite this, the duration and the rate of the development vary between the genotypes, which lead to the formation of fibre with discrete phenotypic differences (Hinchliffe et al., 2010). The value of cotton fibre is mainly dependent on the quality of the fibre defined by length, strength and micronaire. Therefore, it is very important to understand the gene networks and biological processes (BPs) that regulate the fibre development.

In a decade, genomewide analyses of gene expression provided insights into the mechanisms for fibre initiation and development through expressed sequence tags (EST) analysis (Arpat et al., 2004; Haigler et al., 2005; Udall et al., 2006; Wu et al., 2005; Xu et al., 2008), suppression subtractive hybridization (SSH) analysis and macro- or microarray gene expression profiling at a particular stage of fibre development (Alabady et al., 2008; Al-Ghazi et al., 2009; Arpat et al., 2004; Chaudhary et al., 2008; Guo et al., 2007; Hinchliffe et al., 2010; Hovav et al., 2008; Ji et al., 2003; Lee et al., 2006, 2007; Li et al., 2002; Rapp et al., 2010; Shi et al., 2006; Taliercio and Boykin, 2007; Wu et al., 2007; Xu et al., 2007). Earlier, deep-sequencing technology has highlighted the benefits of providing more thorough qualitative and quantitative description of gene expression than the previous microarray-based assay during the early stages (2 to 1 DPA and 2–8 DPA) of cotton fibre development (Wang et al., 2010b).

**Keywords:** cotton, fibre, SRA050044, GSE36228, microarray, transcriptome.
contrasting fibre quality. Therefore, our approach is based on the use of comparative expression profiling of five contrasting genotypes of *Gossypium hirsutum*, viz. JKC 777 and JKC 725 (superior fibre quality) and JKC 703, JKC 737 and JKC 783 (inferior fibre quality) at six developmental stages of fibre (viz. 0, 6, 9, 12, 19 and 25 DPA) using the Affymetrix cotton array and transcriptome sequencing for 25 DPA stage of fibre development in two contrasting genotypes (JKC 777 and JKC 703) using Roche 454 GSFLX genome sequencer.

**Results**

**Genotype grouping based on fibre quality data**

The genotypes used in this investigation, viz. JKC 703, JKC 725, JKC 737, JKC 777 and JKC 783, are being used in various genetic improvement programmes at JK Agri-genetics, Hyderabad, India. The fibre quality parameters revealed that the genotypes JKC 703, JKC 737 and JKC 783 grouped together. This group contains inferior genotypes as their 2.5% fibre length varies from 21.5 to 25.5 mm and fibre strength varies from 21.5 to 23.5 g/tex (Table 1). The genotypes JKC 725 and JKC 777 were grouped as a superior group as their 2.5% fibre span length varies from 30.5 to 32.5 mm and fibre strength varies from 24.5 to 26.0 g/tex, which was significantly higher than the inferior group.

Further, these genotypes were evaluated on the basis of biochemical parameters for fibre length (correlated with dye binding to total fibre unit) and cellulose content (Figure 1). Maximum increase in fibre length was observed during 9–12 DPA (early elongation stage), after which the rate of fibre elongation appears to plateau, but increase in fibre length continued. In the inferior genotypes, fibre elongation occurs at a lower rate than that in the superior genotypes (Figure 1a). JKC 703 had the lowest rate of elongation, while JKC 777 showed the highest rate. Estimation of cellulose content showed that SCW synthesis commenced after 12 DPA until 25 DPA, which overlapped with fibre elongation. Between 19 and 25 DPA is a transition phase where fibre elongation gradually ended and SCW deposition peaked (Figure 1b), in which JKC 703 had the highest rate of cellulose accumulation and JKC 777 the lowest. These genotypes were analysed for genetic relatedness using 1100 AFLP markers, and an UPGMA tree was constructed based on Jaccord’s similarity (Jena et al., 2012) (Figure S1A). The results indicate that the inferior genotypes, viz. JKC 703, JKC 737 and JKC 783, belonged to one group, while the superior genotypes JKC 777 and JKC 725 are more closely related to each other (Figure S1B). Further, to validate the grouping of genotypes obtained from fibre quality data and AFLP analysis, the microarray data of each developmental stage individually were also used to conduct hierarchical clustering, which was consistent with the earlier grouping, except 25 DPA microarray data (Figure S2).

**Microarray data analysis and two-way ANOVA**

The quality of cotton fibre is affected by the genetic make-up of genotypes and also by environmental conditions during growth. A genomewide transcriptome profiling was undertaken using five genotypes of *G. hirsutum* with contrasting fibre quality, viz. superior (JKC 725 and JKC 777) and inferior (JKC 703, JKC 737, JKC 783). The microarray profiling was carried out with Affymetrix cotton chip at six fibre developmental stages (0, 6, 9, 12, 19 and 25 DPA). Thus, a total of 89 microarray experiments were performed for five genotypes and six developmental stages using three biological replicates. The microarray data were analysed by Array-assist software 5.2.2 (Agilent Technologies, Santa Clara, CA) (Padmalatha et al., 2012; Pandey et al., 2013; Ranjan et al., 2012), and differentially expressed genes (DEGs) having statistically significant (*P*-value ≤ 0.05) expression changes (FC ≥ 2.0) were identified by Student’s t-test. Further, to understand the effect of genotypes and developmental stage or both on fibre qualities, we analysed differential gene expression in fibres using Affymetrix cotton genome arrays. The two-way ANOVA (Gao et al., 2007) revealed that 3885 of 21 854 cotton transcripts present on the chip were unique to genotypes and differentially expressed at genotype and temporal level (DPA) with false discovery rate (FDR; 0.1) and

![Figure 1](image-url) Biochemical measurements of fibre lignin (a) and cellulose (secondary cell wall) deposition (b) among various (superior and inferior) genotypes of *Gossypium hirsutum*. 

Table 1 Fibre quality parameters of superior and inferior genotypes taken for study

<table>
<thead>
<tr>
<th>Fibre quality</th>
<th>Superior genotypes</th>
<th>Inferior genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>JKC 725</td>
<td>JKC 777</td>
</tr>
<tr>
<td>2.5% span length (mm)</td>
<td>30.5–32.5</td>
<td>30.5–32.5</td>
</tr>
<tr>
<td>Fibre strength (g/tex)</td>
<td>24.5–26.0</td>
<td>24.5–26.0</td>
</tr>
<tr>
<td>Fineness (micronaire)</td>
<td>3.7–4.0</td>
<td>3.7–4.0</td>
</tr>
</tbody>
</table>
corrected at $P$-value $\leq 0.01$ (Figure 2; Table S1). Majority of DEGs (2721 genes; 71%) were DPA-significant followed by genotype-significant (709 genes; 18%) and interaction-significant (425 genes; 11%) (Figure 2, Table S1). Further to gain an insight into the functional relevance of genes identified by ANOVA, enrichment analysis of BPs for unique DEGs was performed through singular enrichment analysis (SEA) using agrigO tool (http://bioinfo.cau.edu.cn/agriGO/) (Du et al., 2010) for genotypes, DPA and interaction-specific genes (Figure 3a–c, Table S2).

SEA of genotype specific genes showed up-regulation of hormone [auxin, ethylene, jasmonic acid (JA) and salicylic acid], carbohydrate, transport, energy, organelle, fatty acid metabolism along with transcription cofactor and MAP Kinase activity (Figure 3a). Further, DPA-significant genes were primarily involved in four processes including energy- (oxidoreductase, NADPH generation and metabolism, iron–sulphur binding, electron transport chain and pentose phosphate shunt), carbohydrate-, development- and metabolism-related activities (Figure 3b). Interaction significant genes showed up-regulation of processes like cell and growth, peptidase activity, DNA replication, fatty acid biosynthesis and ribosome metabolism (Figure 3c).

**Global transcriptome changes during fibre development between superior and inferior genotypes of *Gossypium hirsutum***

To identify DEGs in superior vs. inferior genotypes, we explored the global transcriptional variation in 8958 genes over a developmental time course of fibre differentiation (0, 6, 9, 12, 19 and 25 DPA) (Figure S3A,B, Tables S3–S5). We noticed that the DEGs were significantly higher during the initiation phase of fibre development (Figure S3A,B), which decrease during fibre elongation stage (9 and 12 DPA) and further increase during SCW stage (19 and 25 DPA). Thus, the results indicate a major transition in the transcriptome during the initiation and SCW stages, while it remains more or less quiescent during the elongation stage. Further, the comparison of superior against inferior genotypes for fibre quality revealed that 10 genes were up-regulated in superior genotypes. These genes were subjected to quantitative real-time RT-PCR between superior (JKC 777) and inferior (JKC 703) genotypes with gene-specific primers (Table S6). The results showed that most of the selected genes were indeed up-regulated in superior genotypes compared with inferior genotypes, thus confirming microarray data (Table 2).

To obtain an overview of gene functional classes that were differentially expressed in the two contrasting genotypes, we conducted over-representation analysis using PageMan, http://mapman.mpimp-golm.mpg.de/pageman/ for comparative analysis of gene ontology (GO) (Usadel et al., 2006). DEGs from each fibre developmental stage (Tables S3 and S4) were used to identify molecular pathway. The significantly over-represented functional groups were identified based on Fisher’s exact test ($P \leq 0.01$). This allowed us to explore the global activation of fibre-specific metabolic pathways during fibre development at two levels: (i) identification of enriched functional categories of metabolic pathways, cellular processes and hormone metabolism during fibre development in the superior and inferior genotypes (Figure 4), and (ii) identification of cluster groups showing distinct time course expression patterns during fibre development (Figures 5 and S4A,B).

**Prolonged activation of lipid metabolism and amino acid synthesis genes at 19 DPA in superior genotypes may result in longer fibres**

PageMan analysis of DPA-specific DEGs showed enriched pathways related to amino acid and lipid activity in superior genotypes (Figure 4). In our study, the phospholipase was found to be up-regulated at 19 DPA in the superior genotypes, which plays a role in membrane biogenesis in plants (Meijer and Munnik, 2003; Ryu, 2004). It was reported that patatin-related lysophospholipase play a role in lipid metabolism, leading to alteration in cellulosic content and cell elongation in *Arabidopsis* (Li et al., 2011). A different member of the patatin gene family encoding PATATIN-LIKE PROTEIN 5 (PLP5; fold change, 15.25) and PLP6 (fold change, 2.16) were significantly over-represented at 19 DPA (Table S3). Along with the lysophospholipase activity, we observed strong expression of monogalactosyldiacylglycerol (MGDG) synthase and UDP sulfoquinose synthase, which are known to play an important role in pollen tube growth in *Arabidopsis* (Figure 4) (Kobayashi et al., 2004). The presence of these activities at this stage of fibre development may result in longer fibres in the superior genotypes.

Recently (Naoumkina et al., 2013) in *Li2* mutant, the levels of nitrogen-containing amino acids such as glutamate, glutamine, aspartate and asparagines were found to be significantly perturbed, emphasizing the role of nitrogen metabolism and asparagine synthase in cotton fibre development. Aspartate family pathway leads to synthesis of essential amino acids lysine, threonine, methionine and isoleucine (Nakamura et al., 2009). These amino acids may serve as precursors for energy generation via glycolysis, amino acid metabolism (synthesis and catabolism) and the TCA cycle (Galili, 1995, 2011). Therefore, the enrichment of amino acid anabolic and catabolic process within fibres of superior genotypes may result in a better source of energy production at this transition stage of peak elongation to SCW, where bulk of energy needs to be diverted for biosynthesis of the SCW.

**Brassinosteroid and PME biosynthetic pathway alterations along with cell wall enzymes during the switch from elongation to SCW in fibre cell of the superior genotypes**

PageMan analysis of six selected clusters (Figure 5a,b) showed that BR hormone expression was more prominent at 9–12 DPA in
Figure 3  Singular enrichment analysis for (a) genotype-, (b) DPA- and (c) interaction-significant genes from two-way ANOVA study.
the superior genotypes compared with the inferior genotypes, and the interaction between hormones leading to cell growth and proliferation has already been reported (Hardtke et al., 2007) (Figure 5c). It was found that BR can crosstalk with numerous other hormones in regulating many developmental processes in plants (Zhang et al., 2009). In the case of cotton fibre cells, BR promotes elongation, which may be a factor in determining their final length (Galili, 2011). A transcript, GhBES1, was found to be highly expressed at 25 DPA (fold change, 7.5) in superior genotype (Table S3). GhBES1 (BRI1-EMS-SUPPRESSOR 1) transcript is known to be a positive regulator of BR signalling, which promotes stem elongation in Arabidopsis (Yin et al., 2002). Along with GhBES1, another gene GhDET2 (DE-ETIOLATED 2) encoding sterol 5-α reductase was also significantly up-regulated at the 19

Table 2 Real-time RT-PCR validation of candidate genes identified through microarray analysis in superior (JKC 777) vs. inferior (JKC 703) genotype

<table>
<thead>
<tr>
<th>Serial no.</th>
<th>Gene ID</th>
<th>Gene annotation</th>
<th>Stages</th>
<th>Fold up (microarray)</th>
<th>Fold change (RT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Ghi.728.2.S1_at</td>
<td>Fringe-related protein</td>
<td>0 dpa</td>
<td>5.2836</td>
<td>2.5996</td>
</tr>
<tr>
<td>2.</td>
<td>Ghi.8379.1.S1_s_at</td>
<td>Putative small heat-shock protein</td>
<td>0 dpa</td>
<td>5.1887</td>
<td>0.9471</td>
</tr>
<tr>
<td>3.</td>
<td>GhiAffx.26755.1.S1_at</td>
<td>Putative ankyrin-repeat-containing protein</td>
<td>0 dpa</td>
<td>3.2670</td>
<td>70.7259</td>
</tr>
<tr>
<td>4.</td>
<td>GhiAffx.53241.1.S1_at</td>
<td>Disease-resistant responsive (dirigent-like protein) family protein</td>
<td>0 dpa</td>
<td>3.0587</td>
<td>111.8173</td>
</tr>
<tr>
<td>5.</td>
<td>GhiAffx.61776.1.S1_at</td>
<td>Arabinogalactan protein AGP10</td>
<td>0 dpa</td>
<td>2.8385</td>
<td>120.2589</td>
</tr>
<tr>
<td>6.</td>
<td>GhiAffx.24987.1.S1_at</td>
<td>Putative peroxidase ATP4a</td>
<td>6 dpa</td>
<td>2.2655</td>
<td>0.5850</td>
</tr>
<tr>
<td>8.</td>
<td>Ghi.1362.1.S1_at</td>
<td>BEL-like homeodomain1</td>
<td>12 dpa</td>
<td>2.6527</td>
<td>5.2355</td>
</tr>
</tbody>
</table>
DPA stage. Previously, it has been shown that GhDET2 plays a crucial role in the fibre cell initiation and elongation, suggesting that the interaction between factors GhBES1 and GhDET2 may regulate fibre quality in the superior genotypes (Luo et al., 2007).

In inferior genotypes at the 12–19 DPA stage of fibre development, significant level of expression of transcripts encoding shaggy-like kinase (GhSK3: fold change, 5.60; and GhSK4: fold change, 4.03) was detected (Table S4). GhBIN2 is one of the
member of the shaggy-like protein kinase gene family that has been identified as a negative regulator of BR signalling in Arabidopsis (Sun et al., 2005) and may have a similar role in the inferior genotypes during elongation.

Pectin is one of the principal components of the primary plant cell wall. In cotton fibres, pectins constitute about 25% of its cell wall and are de-esterified by the pectin methyl esterase (PME) (Hong et al., 2010). Pectin-modifying enzymes such as polygalacturonases (PGs) or pectinases, pectate lyases (PELs), PMEs and PME inhibitors (PMEIs) were differentially up-regulated during the fibre elongation stage in the superior compared with inferior genotypes (Table S3). It was shown that GhPEL, encoding a pectate lyase, plays an essential role in cell wall loosening by depolymerization of the de-esterified pectin during fibre elongation in cotton (Wang et al., 2010a).

Differential abundance of transcription factors (TFs) at initiation, elongation and SCW synthesis stages of fibre development

We examined the level of differentially accumulated TFs in superior and inferior genotypes during the different fibre developmental stages (Table S7). The results showed that 34 TF genes were differentially accumulated at different developmental stages of cotton fibre (Figure 6). Of the 34 TFs, 12 TFs including bHLH, C2H2-ZINC FINGER, MYB, AP2/EREBP, WRKY, NAC, SBP, HSF, HOMEOBOX, bZIP, C2C2-GATA and C3H were differentially accumulated in comparison with the remaining 22 TFs at initiation, elongation and the SCW biosynthesis stages in both superior and inferior genotypes (Figure 6).
Fibre development.

...and few of them have been shown earlier in the regulation of expression level of TFs at different fibre developmental stages. Thus, these results show the contrasting pattern of accumulated only in superior genotypes at initiation and elongation (Wang et al., 2013). The AP2/EREBP, ARF and GRF TFs were found in inferior genotypes. A number of MYB TFs were found to be dominant at initiation (MYB3, MYB109, MYB26, MYB73 and MYB4), elongation (MYB106, MYB42) and SCW (MYB60, MYB82) stages in superior genotypes compared with inferior genotypes, which were reported earlier as a key player for cotton fibre initiation and elongation (Wang et al., 2004). The SVP box TFs were uniquely accumulated only in superior genotypes at initiation and elongation stages. Thus, these results show the contrasting pattern of expression level of TFs at different fibre developmental stages, and few of them have been shown earlier in the regulation of fibre development.

Potential genes identified by microarray study belong to QTL hot spots reported for fibre traits

In case of cotton, meta-QTL analysis of fibre traits resulted in the identification of QTL hot spots (Rong et al., 2007), and recently, this information is being used to assign genes related to fibre trait to hot spot region (Paterson et al., 2012) as the draft D-genome sequence is now available (http://www.cottongen.org/species/Gossypium_raimondii/bgi-cgp_genome_v1.0). We utilized this QTL hot spot information available (Paterson et al., 2012; Rong et al., 2007) for mapping the potential factors identified during our microarray analysis of contrasting genotypes. Our microarray study identified a total of 1684 genes (Table S8) expressing differentially between superior and inferior genotypes during six developmental stages (Figure S3). We mapped these 1684 genes to the draft D-genome using BLAST. Our analysis indicated that 600 of total 1684 genes (~35%) were actually mapped in the QTL hot spot regions reported (Table S9). Interestingly, maximum of 138 of 600 genes (23%) were mapped on the QTL hot spot on chromosome 9, which was in accordance with the earlier report (Paterson et al., 2012). Thus, this analysis provides evidence that the genes reported by us for fibre trait have functional relevance.

Transcriptome assembly and annotation

To obtain an overview of the gene expression profile of later stage (25 DPA) of fibre development, cDNA samples were sequenced using 454 GSFLX genome sequencer from superior (UIC 777) and inferior (UIC 703) genotypes. The digital gene expression analysis was carried out by tagging reads from both the genotypes and then pooling for assembly on gsAssembler (Newbler v2.3; Life Sciences Corp. A Roche company, Indianapolis, IN), generating 17 900 contigs and 53 983 singletons (Table 3). In the case of superior genotype, after filtration of low-quality reads from 488 135 reads (~161.4 Mb), 429 408 high-quality reads were obtained resulting in 16 457 contigs and 45 023 singletons (Table 3). The average contig size was 860 bp with the longest contig in inferior genotype of 3560 bp, whereas superior genotype contained an average contig size of 823 bp with the longest contig of 4991 bp (Figure S5A,C). Unigenes from both libraries were annotated against the ‘nr’ database, resulting in 40 134 hits for inferior genotype and 36 844 hits for superior genotype transcripts (Table S10). There were 7267 unigenes in case of inferior genotype and 17 342 unigenes in case of superior genotype, which were found to be novel (Table 4). The unigenes from both the genotypes were further annotated against the protein database of G. raimondii (Table S11). A total of 33 359 unigenes were annotated in inferior genotype and 37 017 unigenes were annotated in superior genotype, of which 2155 and 2076 unigenes were novel (Table 4). The unigenes from both the genotypes were significant, respectively (Table S12). Further, the unigenes were also annotated against TAIR10 database (Table 4).

Digital gene expression analysis of transcriptome data

The digital gene expression analysis was carried out by tagging reads from both the genotypes and then pooling for assembly on gsAssembler (Newbler v2.3; Life Sciences Corp. A Roche Company, Branford, CT). The genes with fold change ≥2.0 and fold change ≤−2.0 were used for digital gene expression analysis. The transcript length ranges between 100 and 5526 (Figure S5). This resulted in 376 genes up-regulated and 657 genes down-regulated in superior compared with inferior genotypes (Table S13). Further, GO enrichment analysis was performed on these genes to identify BPs, molecular function (MF) and cellular component (CC) (Figure 7; Table S14).

We observed several biologically important processes in the superior genotypes at 25 DPA, activities related to fatty acid biosynthesis, response to oxidative stress (ROS activity) along with ribosome biogenesis and protein transport activity (Figure 7b),
**Table 3** Summary of de novo assembly and merged assembly of 25 DPA transcriptome from fibres of JKC 703 and JKC 777 using Newbler v 2.5.3

<table>
<thead>
<tr>
<th>Parameters</th>
<th>JKC 703</th>
<th>JKC 777</th>
<th>Merged assembly of both genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total reads generated</td>
<td>547 947</td>
<td>488 135</td>
<td>1 036 067</td>
</tr>
<tr>
<td>Total bases generated (Mb)</td>
<td>168.8</td>
<td>161.4</td>
<td>329.3</td>
</tr>
<tr>
<td>Average read size (bp)</td>
<td>307</td>
<td>330</td>
<td>318</td>
</tr>
<tr>
<td>High-quality reads used in assembly</td>
<td>475 658</td>
<td>429 408</td>
<td>942 559</td>
</tr>
<tr>
<td>All contigs (&gt;100 bp)</td>
<td>17 900</td>
<td>16 457</td>
<td>21 308</td>
</tr>
<tr>
<td>Singletons</td>
<td>53 983</td>
<td>45 023</td>
<td>81 120</td>
</tr>
<tr>
<td>Total bases after assembly (Mb)</td>
<td>102.8</td>
<td>92.2</td>
<td>145.9</td>
</tr>
<tr>
<td>Large contigs (&gt;500 bp)</td>
<td>8752</td>
<td>8153</td>
<td>12 947</td>
</tr>
<tr>
<td>Largest contig size (bp)</td>
<td>3560</td>
<td>4991</td>
<td>5008</td>
</tr>
<tr>
<td>Average contig size (bp)</td>
<td>860</td>
<td>823</td>
<td>936</td>
</tr>
<tr>
<td>N50 contig size (bp)</td>
<td>893</td>
<td>838</td>
<td>987</td>
</tr>
<tr>
<td>Aligned reads (%)</td>
<td>86.8</td>
<td>88.0</td>
<td>86.08</td>
</tr>
<tr>
<td>Aligned bases (%)</td>
<td>86.4</td>
<td>87.4</td>
<td>85.0</td>
</tr>
<tr>
<td>Inferred read error (%)</td>
<td>2.0</td>
<td>2.0</td>
<td>1.8</td>
</tr>
<tr>
<td>Q40 plus bases (%)</td>
<td>94.0</td>
<td>94.1</td>
<td>95.5</td>
</tr>
</tbody>
</table>

*Total of contigs and singletons generated after assembly.

**Table 4** Summary of annotation of unigenes from JKC 703, JKC 777 and merged assembly of 25 DPA fibre transcriptome

<table>
<thead>
<tr>
<th>Parameters</th>
<th>JKC 703</th>
<th>JKC 777</th>
<th>Merged assembly of both genotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total unigenes*</td>
<td>71 883</td>
<td>61 480</td>
<td>102 428</td>
</tr>
<tr>
<td>Hits in ‘NCBI nr’ database</td>
<td>40 134</td>
<td>36 844</td>
<td>56 390</td>
</tr>
<tr>
<td>Hits in ‘tair10’ database</td>
<td>33 359</td>
<td>37 017</td>
<td>51 120</td>
</tr>
<tr>
<td>Hits in ESTScan</td>
<td>44 592</td>
<td>40 570</td>
<td>20 852</td>
</tr>
<tr>
<td>Differential unigenes</td>
<td>2156</td>
<td>2076</td>
<td>–</td>
</tr>
</tbody>
</table>

**Discussion**

In spite of plethora of information available on gene expression profiling during distinct phases of fibre development, how the fibre quality traits are being influenced by genetic background remains largely unexplored. Hence, our present investigation targets at generating knowledge about gene networks and metabolic processes that may influence fibre quality traits in contrasting genotypes of *G. hirsutum*. Such knowledge will be extremely useful for developing future cotton varieties with superior fibre traits.

Our microarray analysis resulted in the identification of several DEGs between superior and inferior genotypes (Figure 2; Table S1). These 3885 DEGs identified by two-way ANOVA were divided into three groups: genotype-, DPA- and interaction-significant (Figure 3). It was interesting to note the enrichment of plant hormones in genotype-significant processes (Figure 3a) as several hormones including auxin (Zhang et al., 2011), ethylene (Qin et al., 2007), BR (Sun et al., 2005) and JA (Tan et al., 2012) have been reported to play an important role in fibre development. Therefore, our results probably indicate that these phytohormones are perceived differently in different genotypes. Further, the enrichment of proline-rich family protein in DPA-significant genes (Figure 3b) was interesting as these proteins are known to play a role in development and SCW formation of cotton fibre (John and Keller, 1995). Recently, it was reported that the transcripts involved in lipid metabolism, carbohydrate metabolism, electron transport system, cellulose synthase and sucrose synthase were down-regulated from the 5–20 DPA stage in *fl* mutant (Padmalatha et al., 2012), thus indicating the importance of these activities in fibre development; interestingly, we identified that the expression of many of these genes was up-regulated in superior genotypes compared with inferior genotypes at elongation and SCW stages (Figures 4 and 5c). Similar correlations were also observed between our study and Zhang et al. (2013), who reported that during fibre cell elongation, proteins related to carbohydrate and protein metabolism, transport and redox homeostasis were the most abundant, and glycolysis was found to be the most important process.

Our present study revealed differential abundance of TFs in superior and inferior genotypes at different stages (i.e. initiation, elongation and SCW) of fibre development (Figure 6). Superior genotypes showed higher abundance of *C2H2, bHLH and NAC* TFs and inferior genotypes, *AP2/EREBP, ARF* and *GRF* TFs. Several new TFs identified in this study may be excellent candidates for future functional validation in cotton.

In our study, 35% (600) of total differential genes (1684) were mapped on the QTL hot spot regions (Table S8) reported earlier by meta-QTL analysis (Lacape et al., 2010; Rong et al., 2007). The QTL hot spots were reported to affect the multiple fibre traits in cotton. In our study, maximum numbers of genes (23%) were mapped on QTL hot spot present on chromosome 9. The QTL hot spot region on the chromosome 9 was earlier shown to affect length, uniformity and short fibre content and also fibre elongation and fineness (Paterson et al., 2012). Thus, results indicate the functional relevance of the selected genes based on our global expression profiling of the important fibre traits in cotton. These genes can be potential targets for further validation using transgenic or molecular-breeding-based approach.

We validated expression of six DEGs ([GhSUS4 (SUCROSE SYNTHASE 4)], [GhCSLA03 (CELLULOSE SYNTHASE-LIKE A3)],...
GhKCS6 (3-KETOACYL-COA SYNTHASE 6), GhMYB25, GhMYB109 and GhF3′H (Flavanone 3′ hydroxylase) using quantitative RT-PCR in superior genotype JKC777 and inferior genotype JKC703. The microarray data and quantitative RT-PCR suggest that the expression of selected genes (GhSUS4, GhCSLA03, GhKCS6, GhMYB25, GhMYB109 and GhF3H) showed higher expression in superior genotype JKC777 (Figure S8A–F). GhSUS4 and GhCSLA03, which play a key role in secondary cell wall biosynthesis and fibre development (Li et al., 2013; Xu et al., 2012b), expression of GhSUS4 at initiation and elongation (0, 6 and 9 DPA) and GhCSLA03 at all DPAs (0–25 DPA) were higher in JKC777 as compared to JKC703 (Figure S8A,B). GhKCS6, which is known to play a key role in very-long-chain fatty acid biosynthesis (VLCFAs) and promote cotton fibre elongation (Qin et al., 2007), expression of GhKCS6 was higher from initiation to SCW (0–19 DPA) in JKC777 as compared to JKC703 (Figure S8C). Further, GhMYB25, whose silencing produced fibreless seeds (Walford et al., 2011), and GhMYB109, which specifically expressed in cotton fibre (Suo et al., 2003), expression was higher in JKC777 throughout all developmental stages (0–25 DPA) as compared to JKC703 (Figure S8D,E). Likewise, GhF3′H, whose silencing suppressed fibre development (Tan et al., 2013), was overexpressed.

Figure 7 Modular enrichment analyses for finding significant concurrent annotations: a, b, c for superior and d, e, f for inferior genotype. (a) Enriched biological process (BP), (b) enriched molecular function (MF) and (c) enriched cellular component (CC) in superior genotypes; (d) enriched BP, (e) enriched MF and (f) enriched CC in inferior genotypes.
throughout all the developmental stages (0–25 DPA) in JKC777 as compared to JKC703 (Figure S8F). Thus, genes that were reported to be playing an important role in fibre development showed significantly higher level of expression in superior genotype JKC777 as compared to inferior genotype JKC703. The higher level of expression of these genes may collectively contribute to superior fibre quality in superior genotypes.

Earlier reports on transcriptome analysis of fibre development were mainly focused on early stages of development (Ruan et al., 2009; Wang et al., 2010b; Wu and Poethig, 2006). Our transcriptome sequencing study enumerates the expression changes in the later stages (covering SCW) of cotton fibre development (i.e. 25 DPA), which is important. The quantitative analysis of transcriptome resulted in the identification of genes related to BPs, such as ribosome biogenesis and fatty acid biosynthesis, and response to oxidative stress was significantly up-regulated in the superior genotypes, and those related to proteolysis and response to abscisic acid were significantly up-regulated in the inferior genotypes (Figure 7a,d). The genes that come in MF category belong to structural constituent of ribosome and protein transporter activity specifically up-regulated in superior genotype in comparison with inferior, whereas many other genes conferring aspartic-type endopeptidase activity were more significant in inferior genotype. The CC category also shows good correlation with the above-described functional process (Figure 7c,f). The transcriptome data also revealed up-regulation of genes such as vacuolar ATPase subunit F protein, expansin A4 (EXP A4), lipid transfer protein (LTP), auxin-responsive transcription factor 10 (ARF10), ethylene response factor 1 (ERF1), gibberellin-regulated family protein, ankyrin-repeat-containing protein (AKR) and WD40 repeat family protein (Table S13) in superior genotype (JKC777) as compared to inferior genotype (JKC703). V-ATPase gene has been shown to accumulate during fibre expansion period (Basra and Malik, 1984), and similarly, genes encoding expansin (EXP) and LTP, which are known to function by loosening or cleaving bonds in the cell wall and thus allowing cell expansion (Cosgrove, 1997; Fry, 1995), were overexpressed in superior genotype. Further, GmEXP1 also reported to interact with GhRDL1 and promote boll number and lint yield in transgenic cotton (Xu et al., 2012a). Major plant hormone-responsive genes ERF1 and ARF10 and gibberellin-regulated family protein, which respond to ethylene (Qin et al., 2007), auxin (Zhang et al., 2011) and GA (Xiao et al., 2010), which play a role in cotton fibre development, were also up-

<table>
<thead>
<tr>
<th>Id</th>
<th>Items</th>
<th>Reference support</th>
<th>Reference size</th>
<th>Hyp</th>
<th>Hyp_c</th>
<th>Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Kegg:03010 Ribosome</td>
<td>16</td>
<td>347</td>
<td>192</td>
<td>33 602</td>
<td>1.97E-10 1.30E-08 AT5G58420, AT4G36130, AT5G60670, AT3G06680, AT4G39200, AT5G56710, AT3G16080, AT1G74060, AT1G27400, AT4G14320, AT3G49910, AT2G30260, AT2G40590, AT1G74270, AT5G48760, AT3G52580</td>
</tr>
<tr>
<td>10</td>
<td>Kegg:00564 Glycerophospholipid metabolism</td>
<td>4</td>
<td>347</td>
<td>61</td>
<td>33 602</td>
<td>0.003671 0.04385 AT4G22340, AT4G19500, AT2G42010, AT4G35790</td>
</tr>
<tr>
<td>20</td>
<td>Kegg:03050 Proteasome</td>
<td>4</td>
<td>347</td>
<td>54</td>
<td>33 602</td>
<td>0.002354 0.03884 AT5G09900, AT5G66140, AT3G60820, AT3G53970</td>
</tr>
<tr>
<td>28</td>
<td>Kegg:03040 Spliceosome</td>
<td>8</td>
<td>347</td>
<td>108</td>
<td>33 602</td>
<td>1.71E-05 0.000565 AT2G32600, AT5G20160, AT1G21190, AT2G38730, AT5G08290, AT3G62840, AT4G03120, AT5G02530</td>
</tr>
<tr>
<td>54</td>
<td>Kegg:00410 beta-Alanine metabolism</td>
<td>3</td>
<td>347</td>
<td>30</td>
<td>33 602</td>
<td>0.003605 0.047585 AT3G48000, AT4G31810, AT4G16210</td>
</tr>
<tr>
<td>55</td>
<td>Kegg:00640 Propanoate metabolism</td>
<td>4</td>
<td>347</td>
<td>29</td>
<td>33 602</td>
<td>0.000216 0.004762 AT3G48000, AT4G31810, AT5G15530, AT4G16210</td>
</tr>
<tr>
<td>17</td>
<td>Kegg:01110 Biosynthesis of secondary metabolites</td>
<td>26</td>
<td>603</td>
<td>678</td>
<td>33 602</td>
<td>0.000278 0.011544 AT2G07050, AT1G54220, AT5G08280, AT5G11380, AT5G59900, AT3G26290, AT2G30490, AT5G53970, AT1G09830, AT2G42940, AT5G63890, AT1G02640, AT3G48000, AT1G65600, AT1G74090, AT1G23190, AT5G47340, AT4G36220, AT5G15530, AT1G79530, AT1G20050, AT5G12400, AT2G37040, AT4G38220, AT5G67400, AT3G41930</td>
</tr>
<tr>
<td>35</td>
<td>Kegg:03015 mRNA surveillance pathway</td>
<td>7</td>
<td>603</td>
<td>74</td>
<td>33 602</td>
<td>0.000368 0.010193 AT1G17980, AT5G19400, AT3G07810, AT1G18070, AT4G15415, AT1G17770, AT5G25510</td>
</tr>
<tr>
<td>67</td>
<td>Kegg:00250 Alanine, aspartate and glutamate metabolism</td>
<td>6</td>
<td>603</td>
<td>43</td>
<td>33 602</td>
<td>0.000113 0.009375 AT5G56360, AT1G29900, AT1G72330, AT1G79440, AT3G47340, AT1G70580</td>
</tr>
</tbody>
</table>

Table 5 Summary of KEGG pathways for differential gene from JKC 703 and JKC 777 of 25 DPA fibre transcriptome
regulated in superior genotype. Ankyrin-repeat-containing protein (AKR) was reported to function as molecular chaperons for ascorbate peroxidase (APX) (Shen et al., 2010); APX was reported to mediate hydrogen peroxide homeostasis in cotton fibre (Li et al., 2007), and the AKR gene was also found to be up-regulated in superior genotypes. Superior genotypes also exhibited up-regulation of WD40 repeat family protein, which plays a role in determining epidermal trichome cell patterning in Arabidopsis leaves (Ramsay and Glover, 2005). Thus, higher expression of these DEGs in superior genotype JK777 based on transcriptome analysis may also relate to its superior fibre quality as compared to inferior genotype BC70.

**Experimental procedures**

**Plant materials and tissue collection**

The plant materials include two genotypes of superior fibre quality (JKC 777 and JKC 725) and three inferior genotypes (JKC 703, JKC 737 and JKC 783) of *G. hirsutum*. These genotypes were grown at the experimental plot of JK Agri-Genetics Limited, Hyderabad, India. Cotton bolls were harvested at 0, 6, 9, 12, 19 and 25 DPA for the isolation of total RNA from developing fibre cells. In the case of 0 DPA, whole ovule was taken for RNA isolation.

**Lignin test for fibre growth**

To assess the fibre growth, the lignin content was estimated by the method described by (Kwak et al., 2009). Briefly, three ovules from bolls of each developmental stage were boiled in 10 mL milliQ water with a drop of concentrated HCl followed by washing with milliQ and staining in 10 mL of toluidine blue solution (0.018% toluidine blue, 0.016 M Na2HPO4, 0.01 M citric acid, pH 4.5) for 30 s. They were placed in 50 mL of destaining solution (glacial acetic acid–ethanol–water; 10: 95: 5) for 1 h. Absorbance was measured at 624 nm on UV spectrophotometer (Shimadzu, Kyoto, Japan). Three biological replicates for each developmental stage as well as for genotypes were processed.

**Cellulose assay for secondary wall deposition**

To measure the cellulose content in developing fibres, the fibres from three ovules (all the developmental stages except 0 DPA) were dissected and ground in liquid nitrogen. The cellulose content was estimated as per the method described previously (Updegraff, 1969). In brief, the pellet was treated with 3 mL of CH3COOH + HNO3 reagent (150 mL 80% CH3COOH, 15 mL HNO3) for 30 min at 100 °C and centrifuged at 3500 g for 5 min and rinsed with de-ionized water. Then, the pellet was treated with 10 mL of concentrated H2SO4 and incubated for 1 h at room temperature. About 500 mg of pure cellulose was taken as reference to plot a standard graph. A total of 100 mL of sample was reacted with 200 mL of anthrone reagent (0.2% anthrone in conc. H2SO4) in 96-well flat-bottom plates. The absorbance was measured at 620 nm in a plate reader (DynA Quant 200; Hoefer, East Lyme, CT).

**RNA isolation and cotton fibre cDNA microarray hybridization**

The total RNA was extracted from developing fibres using the Spectrum Plant Total RNA Kit (Sigma, St. Louis, MO) according to the manufacturer’s instructions. One microgram of total RNA was used to synthesize double-stranded cDNA with T7 promoter-containing oligo(dT) primer using a GeneChip One-Cycle cDNA Synthesis Kit (Affymetrix, Santa Clara, CA), followed by *in vitro* transcription using a GeneChip IVT Labeling Kit (Affymetrix). cRNA was fragmented for hybridization and incubated at 45 °C for 16 h at 18 g. A summary of the image signal data for every gene interrogated on the array was generated using the Affymetrix statistical MAS 5.0 (GCOS v1.3; Affymetrix, Santa Clara, CA) algorithm. All data are MIAME compliant, and the raw data have been deposited in a MIAME compliant database (e.g. GEO), as detailed on the MGED Society (http://www.mged.org/Workgroups/MIAME/miame.html).

**Statistical analysis**

Images were acquired by scanning the chips with Affymetrix Gene Chip Scanner (Signals ≥ 2000). All the statistical analyses of differential gene expressions were carried out with Array-assist 5.2.2 (Agilent Technologies) (Padmalatha et al., 2012; Pandey et al., 2013; Ranjan et al., 2012) except two-way ANOVA, which was performed using MeV v4.3.01 (TM4, Microarray suit, Boston, MA). RNA-normalized data values for each microarray were log2-transformed and median-centred across all 89 arrays, which were used to calculate Pearson’s correlation between genotypes and their biological replicates. Correlation values for each DPA were used for constructing dendrogram for all five genotypes (Figure S2). Further, log2-transformed data were analysed using (i) Student’s t-test (*P*-value ≤ 0.05) and fold expression changes (FC ≥ 2.0). This produced a total of 36 lists of up- and down-regulated genes across developmental time.

**Cluster analysis**

Chronological data were used for SOM (self-organizing maps) cluster analysis using AutoSOME (http://jimcooperlab.mcdb.ucsb.edu/autosome/). This analysis was performed using a Euclidean distance matrix (*P*-value ≤ 0.01) generating 30 fuzzy (random) clusters (Figure S4A–B) for superior and inferior genotypes. But, three clusters each in both genotypes have common and distinct expression profiles representing initiation-, elongation- and SCW-specific expressions, which were selected for PageMan analysis.

**Genomewide transcription factor identification**

The genomewide TFs were identified in three developmental stages: initiation (0–6 DPA), elongation (9–12 DPA) and SCW (19–25 DPA), within the DEGs using 0 DPA as a control using *Arabidopsis* transcription factor database (http://arabidopsis.med.ohio-state.edu/AtTFDB/).

**Cotton gene chip annotation**

*Arabidopsis* homologues of the cotton probe set were specified using Plexdb tool (http://www.plexdb.org/). Cotton gene chip sequences were also annotated with *G. raimondii* genome (version 2.2.1) using BLASTx with e-value ≤ 10−5 (Table S16).
Quantitative real-time RT-PCR
First-strand cDNA was synthesized from 2 μg DNaseI-treated total RNA using Superscript II Reverse Transcription kit (Invitrogen, Carlsbad, CA) as per the manufacturer’s instructions. The cotton polyubiquitin (GhUB1, accession number in GenBank: EU604080) was used as the reference gene. Quantitative real-time RT-PCR amplification was carried out using the ABI Prism 7000 sequence detection system (Applied Biosystems, Foster City, CA) with Power SYBR green PCR Master Mix (Applied Biosystems) according to the manufacturer’s instructions.

Transcriptome sequencing
Sequencing library preparation and 454 sequencing were done using Roche GSFLX Titanium sequencing kit (Branford, CT). Five microgram of DNaseI-treated total RNA was reverse-transcribed using One-Cycle cDNA Synthesis kit (Affymetrix). These libraries were sequenced on Roche 454 Genome Sequencer (Titanium, GS sequencer v2.5). Signal processing was performed using gsRunBrowser v2.5.3 (Life Science, Roche company, Indianapolis, IN).

De novo assembly, annotation and digital gene expression analysis of transcriptome sequences
All high-quality reads generated from transcriptome sequencing were de novo assembled using gsAssembler (Newbler v2.3) with 40 bases of overlap and 90% of identity. Unigenes (including contigs and singletons) from both the genotypes were queried for their putative function using BLASTx program against the NCBI ‘nr’ protein database (4 June 2012) and Arabidopsis protein database (TAIR10) with e-value ≤ 10^{-5}. For digital gene expression analysis, reads from each library were tagged and pooled to form a large data set, which was assembled using gsAssembler (Newbler v2.3) with 40 bases overlap and 90% identity. From resulting assembly, transcript per million (TPM) was calculated, which was further log2-converted. Fold change was calculated between superior and inferior genotypes. The contigs with fold change ≥ 2.0 or ≤ -2.0 were used for GO analysis using GeneCodis3 (http://genecodis.cnb.csic.es/) (Tabas Madrid et al., 2012), and the GO categories were further divided into MF, BP and CC.

Correlation of transcriptome sequences and microarray data
The unigenes from each 25 DPA libraries were queried against the cotton chip probe set sequences (Affymetrix; http://www.affymetrix.com/support/technical/byproduct.affx?product=cotton) using standalone BLASTn program (e-value ≤ 10^{-5} and alignment length ≥ 100 bases with identity ≥ 90%). The fold change values of the filtered probe sets were used for Spearman’s correlation analysis with the corresponding transcriptome data.

QTL hot spot identification
All the differential genes in each library were mapped according to the G. raimondii jgi genome (http://www.cottongen.org/species/Gossypium_raimondii/bgi-cgp_genome_v1.0) using BLASTn with e-value ≤ 10^{-10}. Then their position was mapped against the QTL hot spot location on each chromosome as mentioned by (Paterson et al., 2012). All those genes were selected that have their start and end position within the QTL hot spot.

Data access
The microarray data are submitted to NCBI GEO (GSE36228). The 454 sequencing reads have been submitted to NCBI SRA (SRA050044), and the accession number for G. hirsutum superior genotype JKC 777 is SRR633517 and for inferior genotype JKC 703 SRR633518.

Acknowledgements
This work was funded by Council of Scientific and Industrial Research (CSIR), India, under NIMITLI (TLP400925) and Institutional Projects (SIP05 and BSC0107). DN and AR thank CSIR, India, for providing fellowship.

Conflict of interest
The authors declare that they have no competing interests.

References


Supporting information

Additional Supporting information may be found in the online version of this article:

**Figure S1** AFLP based genetic relationship of five genotypes of *Gossypium hirsutum* using 16 primer pair combinations of 1100 markers (A) UPGMA tree (B) Jaccard’s similarity.

**Figure S2** Cluster analysis on the basis of microarray data for six DPAs in five genotypes.

**Figure S3** Time point based study for identification of genotype specific fiber differential genes for six developmental time points (i.e. 0, 6, 9, 12, 19 and 25 DPA) based on two strategies (A) superior to inferior comparison (B) chronological comparison.

**Figure S4** (A–B) 30 clusters identified for Superior and inferior genotypes.

**Figure S5** Size distribution of Transcriptome data.

**Figure S6** Spearman correlation plot in between microarray and transcriptome data for 25 DPA.

**Figure S7** Heat map of differentially expressed genes at 25 DPA microarray and 25 DPA transcriptome data of *G. hirsutum* showing similar expression trend.

**Figure S8** Expression of six known and validated fiber related genes (*GhSUS4, GhCSLA03, GhKCS6, GhMYB25, GhMYB109 and GhF3H*) in both superior and inferior genotypes.

**Table S1** Genotypes (sheet 1), interaction (sheet 2) and DPA-significant (sheet 3) genes from two-way ANOVA study.

**Table S2** SEA analyses for genotypes (sheet 1), interaction (sheet 2) and DPA-significant (sheet 3) genes from two-way ANOVA study.

**Table S3** Differentially expressed gene in superior genotypes at their five transition stage, that is, chronological stage (0 vs. 6, 6 vs. 9, 9 vs. 12, 12 vs. 19, 19 vs. 25).

**Table S4** Differentially expressed gene in inferior genotypes at their five transition stage, that is, chronological stage (0 vs. 6, 6 vs. 9, 9 vs. 12, 12 vs. 19 and 19 vs. 25).

**Table S5** Differentially expressed genes (up- and down-regulated) identified in superior genotypes at 0 DPA (SS sheet 1, sheet 2), six DPA (SS sheet 3, sheet 4), nine DPA (SS sheet 5, sheet 6), 12 DPA (SS sheet 7, sheet 8), 19 DPA (SS sheet 9, sheet 10), 25 DPA (SS sheet 11, sheet 12) of fibre development.

**Table S6** Primer sequences of genes used for qRT-PCR validation.

**Table S7** Enriched transcription factors in superior and inferior genotypes at initiation, elongation and SCW (secondary cell wall) stages of fibre development.

**Table S8** Gene ids mapped on QTLs present in different chromosome.

**Table S9** Positions of the mapped differentially expressed genes on QTL.

**Table S10** Annotation detail for contigs and singletons from JK777 (36 843), JK703 (40 134) and merge assembly against ‘nr’ database.

**Table S11** Annotation detail for contigs and singletons from JK777 (36 843), JK703 (40 134) and merge assembly against ‘G. raimondii protein’ database (25 DPA transcriptome).

**Table S12** Annotation detail for contigs and singletons from JK777 (36 843) and JK703 (40 134) against ‘TAIR 10 protein’ database.

**Table S13** Differentially expressed genes identified in JK777 compared with JK 703 through digital gene expression analysis.

**Table S14** GO annotation of differentially expressed genes of JK777 and JK 703 through GeneCodis3 tool.

**Table S15** Spearman’s correlation calculated between Affymetrix microarray and 454 transcriptome sequencing data set of 25 DPA cotton fibre.

**Table S16** Annotation details of Cotton chip sequences against ‘G. raimondii protein’ database.